"I AGREE TO THE TERMS AND CONDITIONS":
legal-lay language complexity in English and Italian

Lucia Busso, Aston Institute for Forensic Linguistics
l.busso@aston.ac.uk

BACKGROUND and MOTIVATION

Legal-lay language (LLL): the language of texts with legal content aimed at non-specialist audiences
→ Increasingly important in our everyday life (Benoliel & Becher 2019) → should be written in "plain language", BUT linguistic and legal literature agrees that LLL often is too complex (Conklin et al. 2019; Frade 2007) → many problems, legislative and linguistic alike (Bhatia 1983; Haapio 2011).
→ First linguistic study that triangulates evidence from corpus-based, computational and experimental methodologies
→ Theoretical framework: Construction grammar (Goldberg 2006, 2019)

RESEARCH QUESTIONS

1. **Is LLL quantitatively different from specialist legal jargon and general domain written language?**
2. **Is LLL more readable and more comprehensible than legal jargon? Do readability measures efficiently capture text comprehensibility?**
3. **What are the principal components of LLL's lexical and syntactical complexity with respect to legal jargon and general-domain written prose?**

DATA AND METHODOLOGY

**CorIELLS (CORpus of Italian and English Legal-Lay textS)** > 1.7M words
Methods:
- ❖ collostructional analysis (Stefanowitsch 2013) + core vocabulary comparison (NGSL, Brezina & Gablasova 2015; NVDB, Chiari & De Mauro, 2016)
- ❖ Comparative analysis with legal language and general domain written corpora
- ❖ Readability metrics (ARI, Coleman-Liau, Flesch-Kincaid) + survey (Eng 50, Ita 103)
- ❖ Principal Component analysis of linguistic complexity measures
- → 4 constructions: MOD, NOM_PP, PART, PASS for all analyses

Document type	Document nº	English	Italian
Bank contracts	15 in total	19%	27%
Utilities contracts	15 per language	25%	23%
EurLex summaries	247 per language	22%	23%
Terms and conditions	45 per language	34%	27%

RESULTS

English			
Constructions	Collexemes	%presence	%absence
MOD	94	75.5%	24.5%
NOM_PP	50	66%	**34%**
PART	94	79.8%	20.2%
PASS	100	77%	23%

Italian			
MOD	147	80.3%	19.7%
NOM_PP	100	68%	**32%**
PART	21	66.7%	**33.3%**
PASS	121	76%	24%
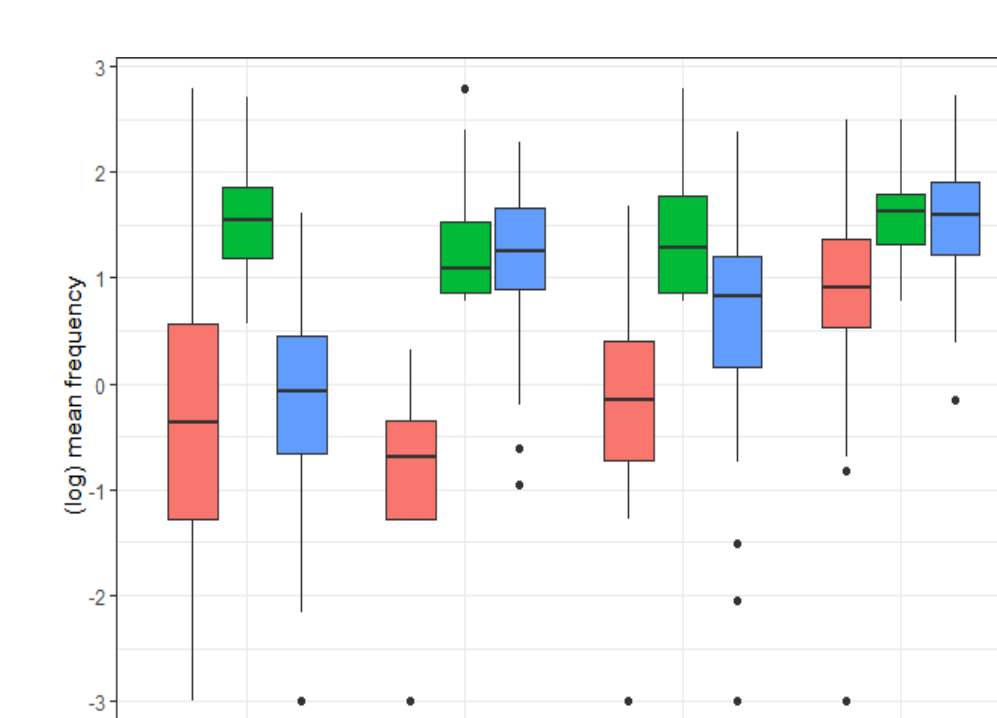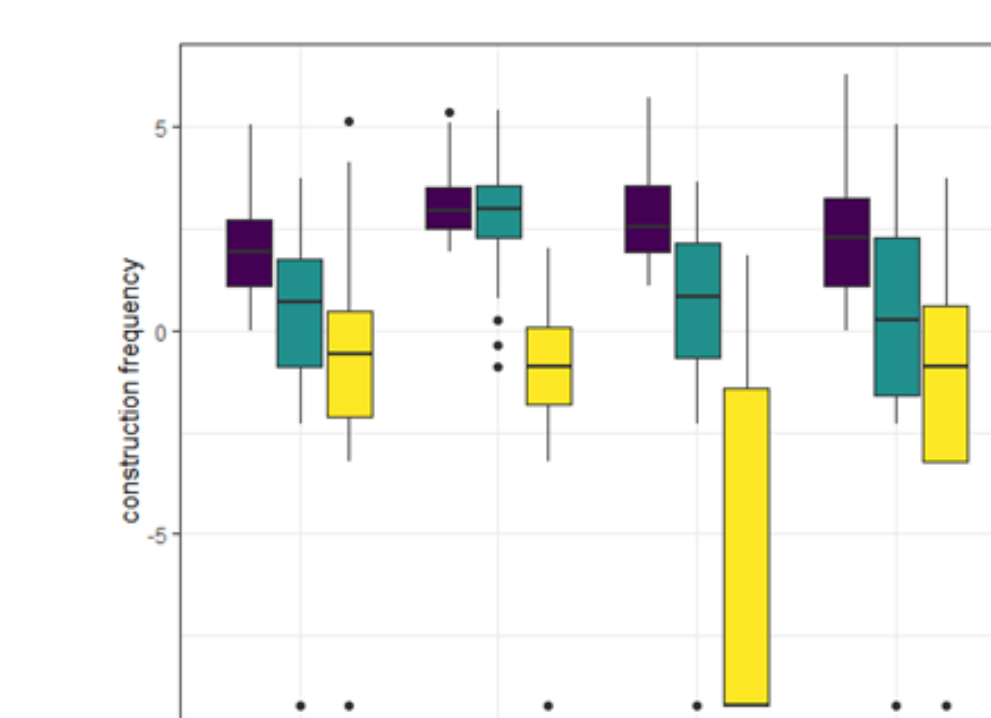
① linear mixed effect modelling confirms that LLL has an intermediate status between specialised legal jargon and general written language → significant differences (p <.0001) and high effect sizes for both languages for all constructions except NOM_PP (p > .05) → the most legal construction??
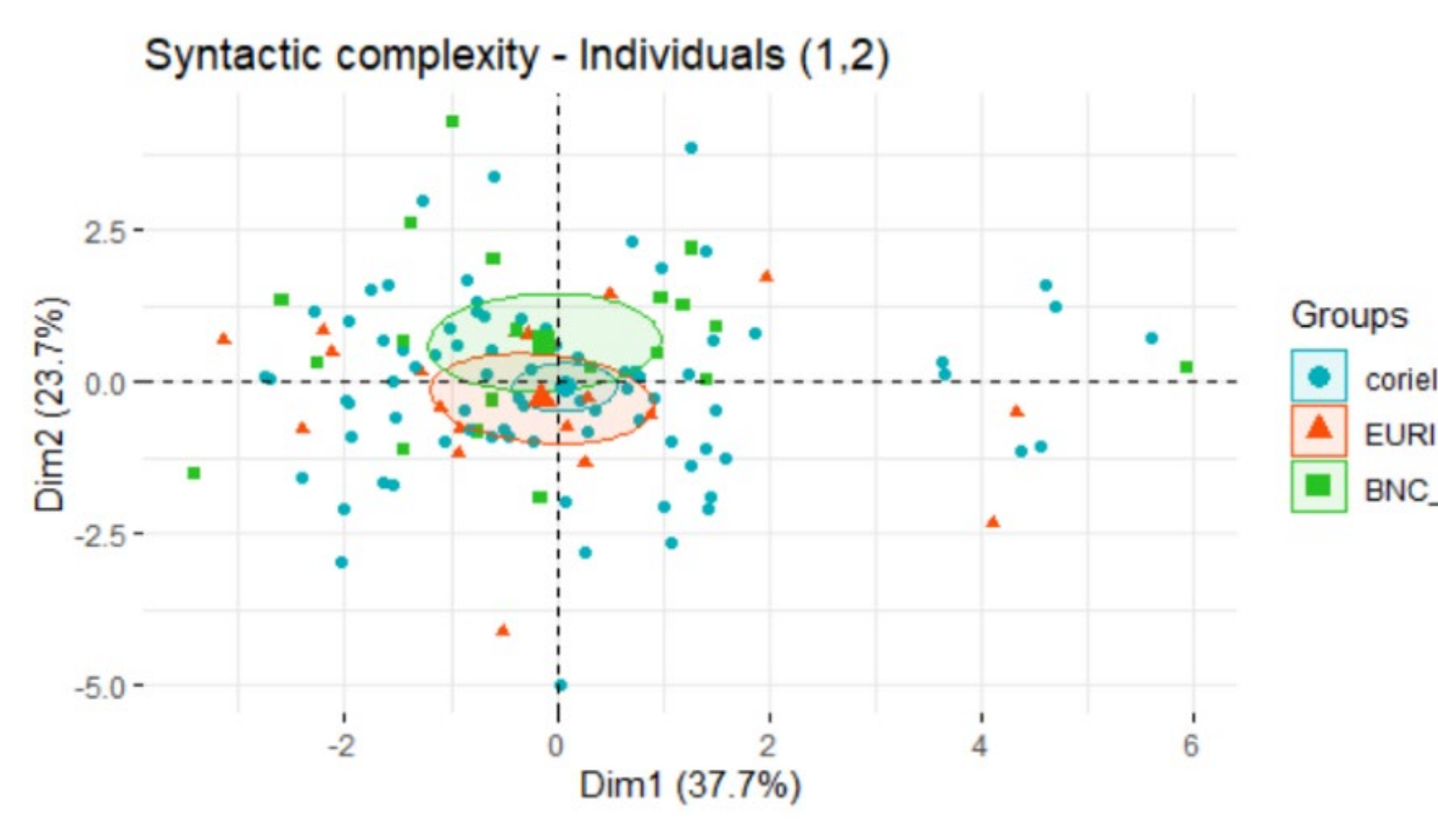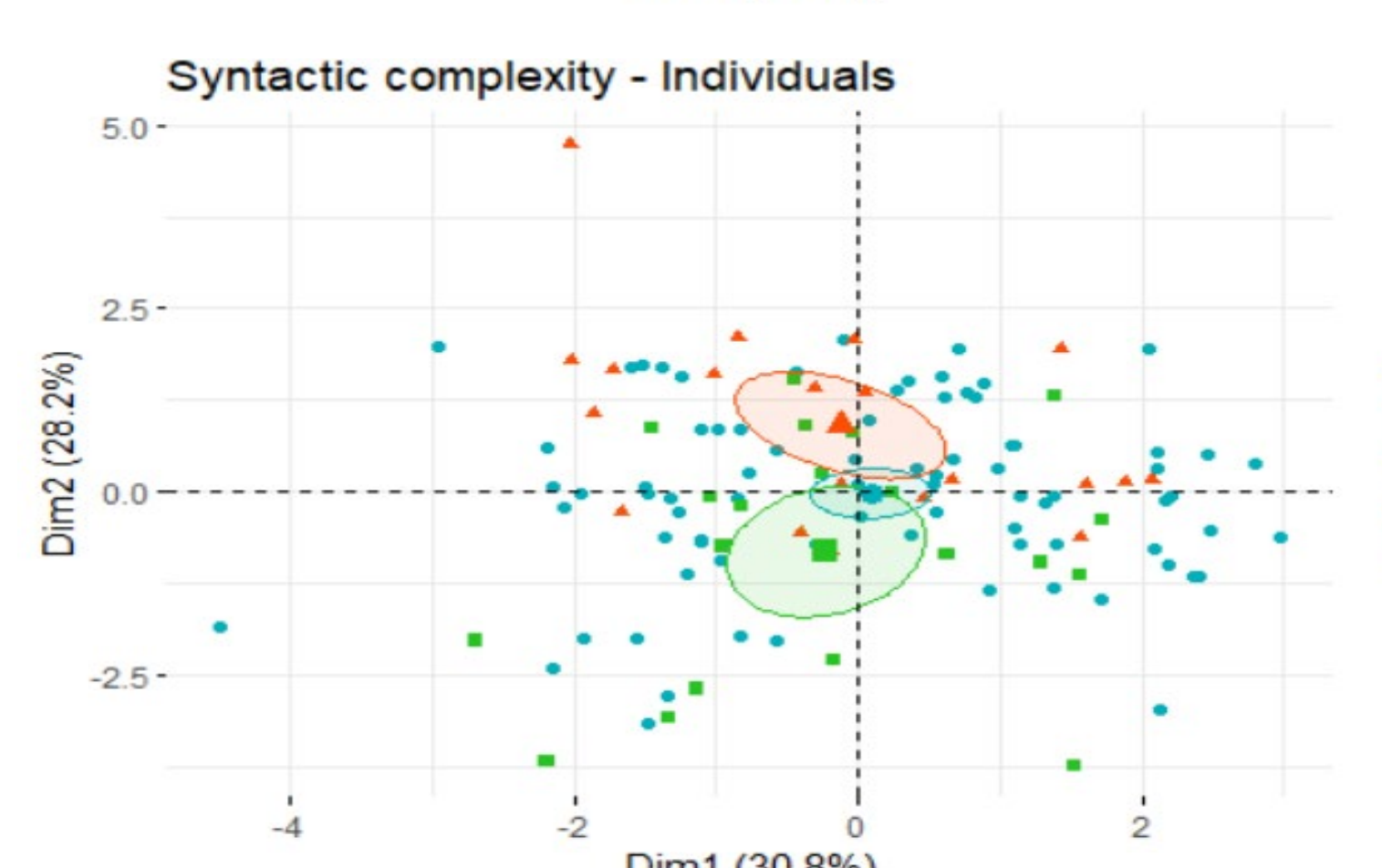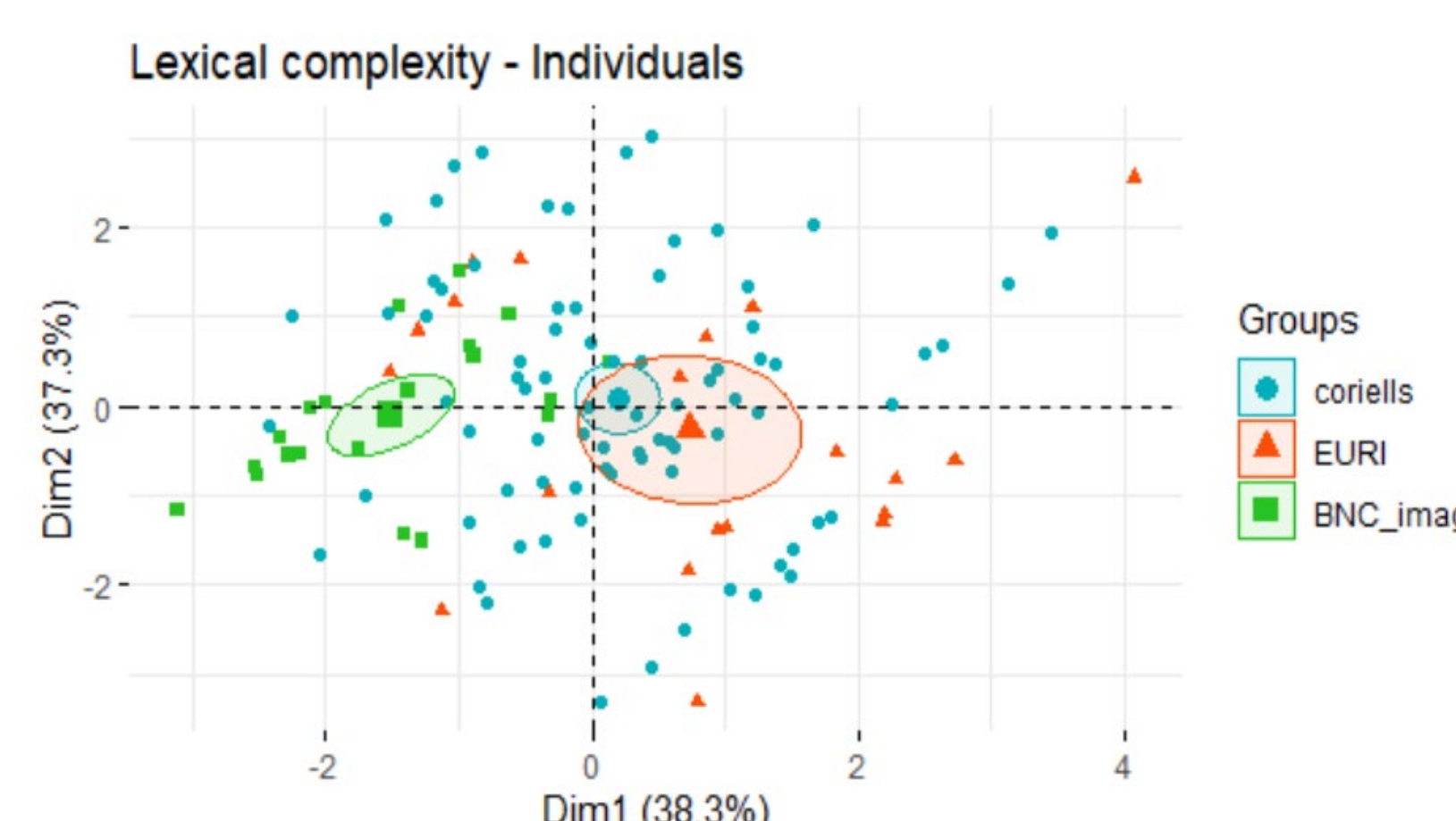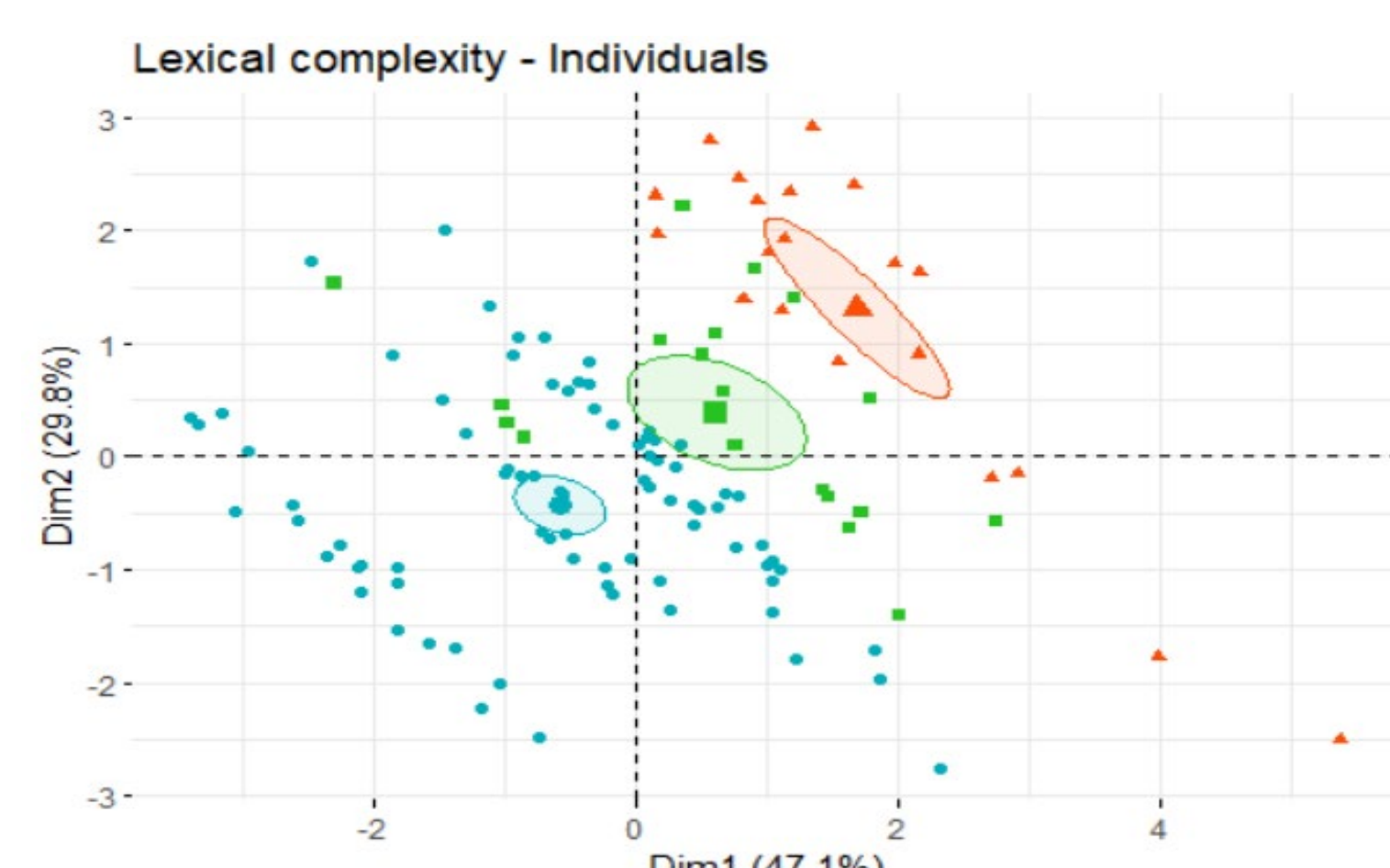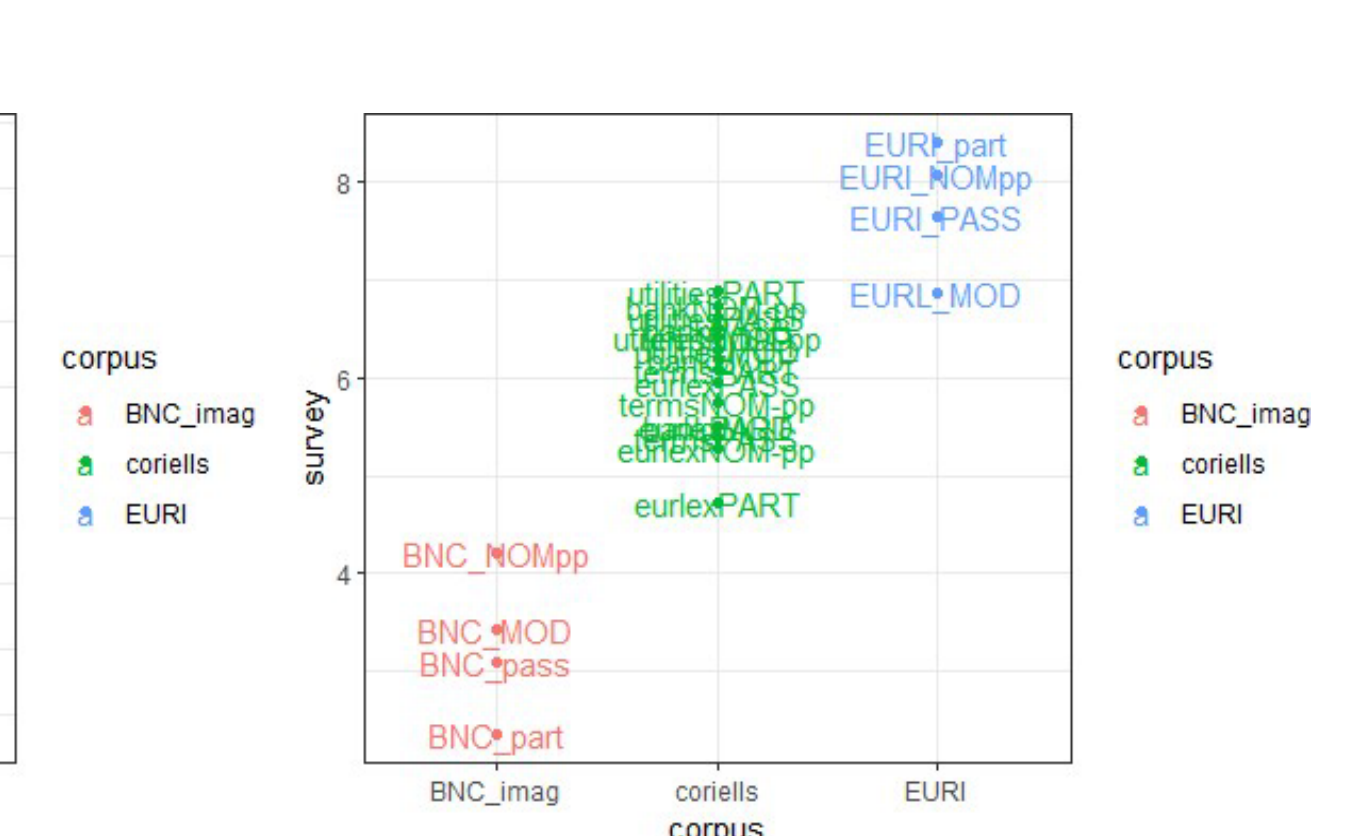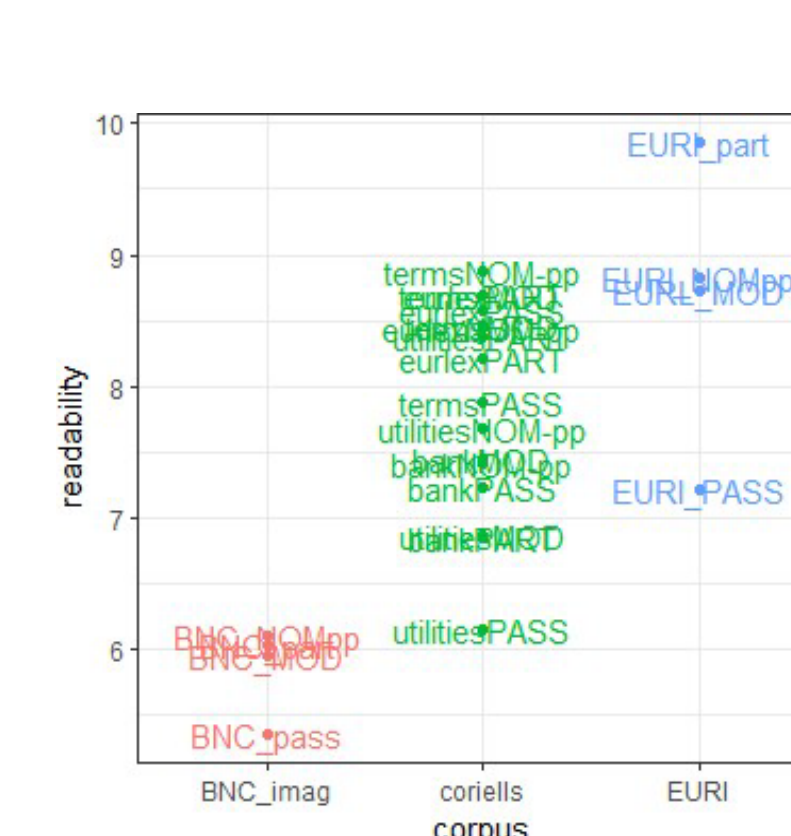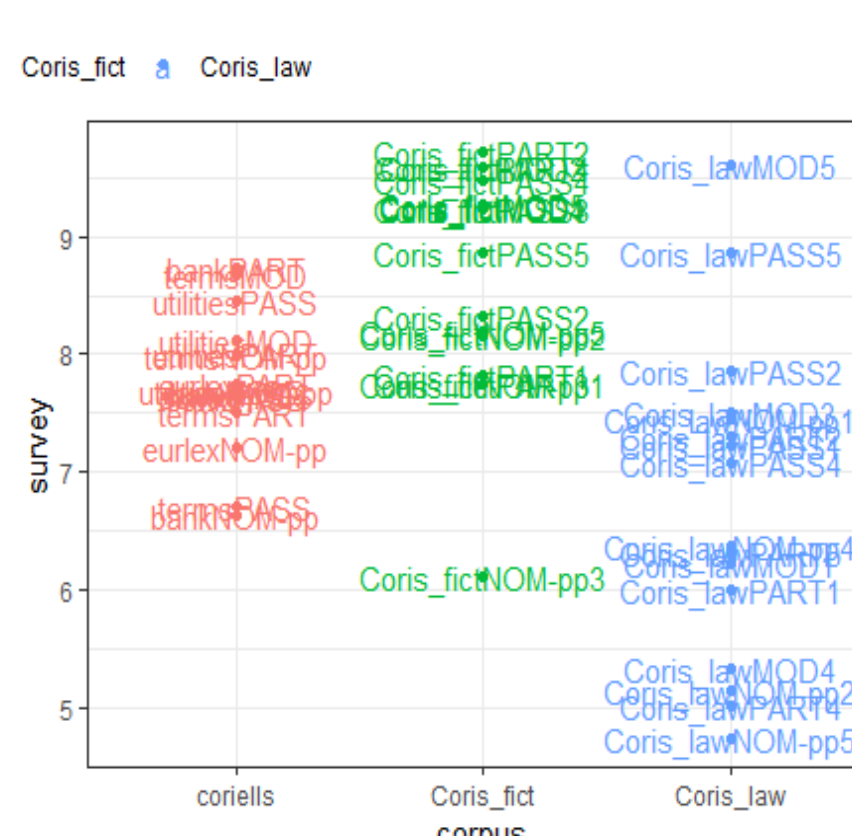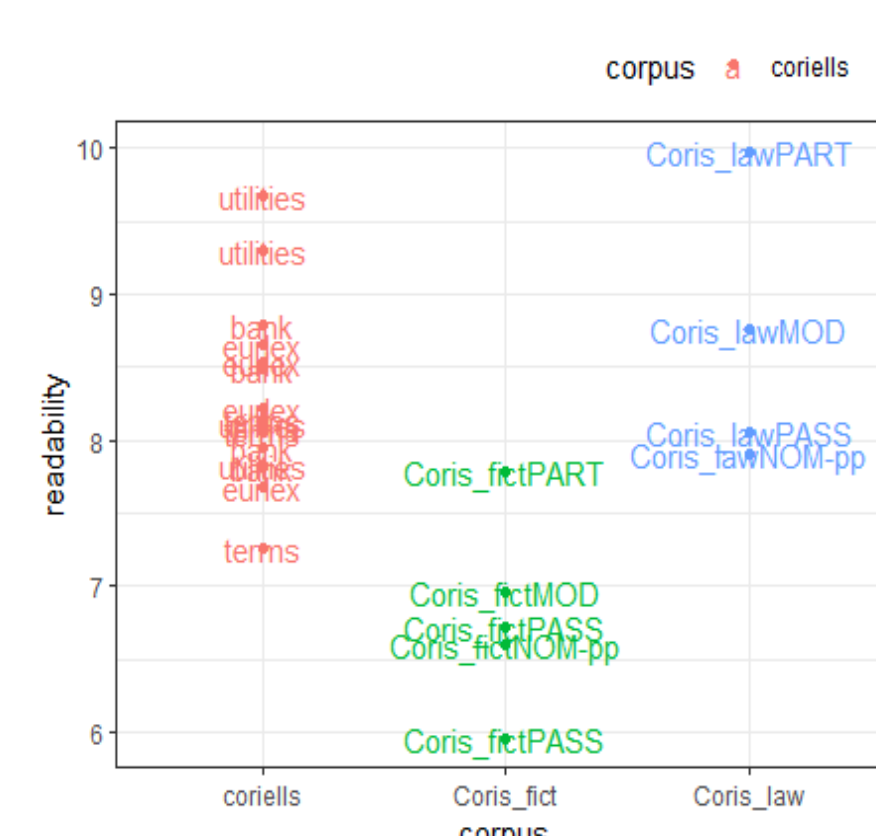
② Readability metrics and speakers' judgments confirm the trend of the «blended» nature of LLL in comprehensibility as well. However: readability metrics underestimate speakers' ability to extract meaning.
Tested with ANOVA (condition: p <.0001, corpora: p <.0001).

Lexical complexity - Individuals

Lexical complexity - Individuals

Syntactic complexity - Individuals

Syntactic complexity - Individuals (1,2)

③ 137 lexical and syntactical dimensions extracted from ProfilingUD (Brunato et al. 2020) and analysed with PCA.

English: LEXICAL: 1) Lexical Diversity, (2) Lexical Density
SYNTACTICAL: (1) Syntactic Structure, (2) Dependency Length
→ Lexically English LLL is similar to legal jargon; syntactically, much overlap

Italian: LEXICAL: 1) Lexical Diversity, (2) Lexical Density.
SYNTACTICAL (1) Dependency structure, (2)Internal constituents' complexity.
→ Perfect mix for morphosyntactic dimensions, idiosyncratic lexically